DEFENCE
SCIENCE & TECHNOLOGY

DSTO

Fitting the Most Likely Curve through Noisy Data

Garry N. Newsam and Nicholas J. Redding

DSTO-RR-0242

Copy No. of 64

DEFENCE
SCIENCE & TECHNOLOGY

Fitting the Most Likely Curve through Noisy Data

Garry N. Newsam and Nicholas J. Redding

Surveillance Systems Division
Electronics and Surveillance Research Laboratory

DSTO–RR–0242

ABSTRACT

At present the preferred method for fitting a general curve through scattered data points in the plane is *orthogonal distance regression*, *i.e.*, by minimising the sum of squares of the distances from each data point to its nearest neighbour on the curve. While generally producing good fits, in theory orthogonal distance regression can be both biased and inconsistent: in practice this manifest itself in overfitting of convex curves or underfitting of corners. The paper postulates this occurs because orthogonal distance regression is based on an incomplete stochastic model of the problem. It therefore presents an extension of the standard model that takes into accounts both the noisy measurement of points on the curve and their underlying distribution along the curve. It then derives the likelihood function of a given curve being observed under this model. Although this cannot be evaluated exactly for anything other than the simplest curves, it lends itself naturally to asymptotic approximation. Orthogonal distance regression corresponds to a first order approximation to the maximum likelihood estimator in this model: the paper also derives a second order approximation, which turns out to be a simple modification of the least squares penalty that includes a contribution from the curvature at the closest point. Analytical and numerical examples are presented to demonstrate the improvement achieved using the higher order estimator.

APPROVED FOR PUBLIC RELEASE

Published by

DSTO Electronics and Surveillance Research Laboratory
PO Box 1500
Edinburgh, South Australia, Australia 5111

Telephone: (08) 8259 5555
Facsimile: (08) 8259 6567

AR No. 012–374
May, 2002

APPROVED FOR PUBLIC RELEASE

Fitting the Most Likely Curve through Noisy Data

EXECUTIVE SUMMARY

Many problems in computer vision and elsewhere centre on finding some member of a family of curves that best fits a given set of points in a plane (or, more generally, on finding a member of an m-dimensional family of surfaces that best fits a set of points in n-dimensional space). At present the preferred method for fitting a general curve through scattered data points in the plane is *orthogonal distance regression*, *i.e.*, by minimising the sum of squares of the distances from each data point to its nearest neighbour on the curve. Orthogonal distance regression is well known to be computationally difficult, and although some good algorithms are now available, it can still be very hard to numerically determine the optimal fit. This paper is not concerned with efficient computation, however. Rather it explores the more fundamental problem of the flaws in the stochastic model on which orthogonal distance regression is based, and develops an improved model and associated distance measures that produce better fits. While generally producing good fits, in theory orthogonal distance regression can be both biased and inconsistent: in practice this manifest itself in overfitting of convex curves or underfitting of corners. The paper postulates this occurs because orthogonal distance regression is based on an incomplete stochastic model of the problem.

The paper stemmed from the authors' practical experience in fitting curves to ionogram data, while the approach taken was inspired by some perceptive remarks by Kanatani on the distinction between geometric and statistical distance measures in fitting curves to data. It was found that curves fitted to these data sets by orthogonal distance regression had a tendency to underfit the nose of the trace (*i.e.*, not to extend as far as the data). Theoretical confirmation of this bias in the fitting procedure would have serious implications for ionogram modelling due to the importance of accurately locating the exact extent of the nose in applications of interest.

This paper therefore presents an extension of the standard orthogonal distance regression model that takes into accounts both the noisy measurement of points on the curve and their underlying distribution along the curve. It then derives the likelihood function of a given curve being observed under this model. Although this cannot be evaluated exactly for anything other than the simplest curves, it lends itself naturally to asymptotic approximation. Orthogonal distance regression corresponds to a first order approximation to the maximum likelihood estimator in this model: the paper also derives a second order approximation, which turns out to be a simple modification of the least squares penalty that includes a contribution from the curvature at the closest point. Analytical and numerical examples are presented to demonstrate the improved fits possible using the higher order estimator, along with an extension to higher dimensional problems involving surface fitting.

Authors

Garry N. Newsam
Surveillance Systems Division

Garry Newsam received a B.Sc. and M.Sc. in mathematics from the University of Canterbury, New Zealand in 1978, and an A.M. and Ph.D. in applied mathematics from Harvard University in 1983. He then worked for a year as a research fellow at Victoria University, New Zealand examining the statistics of energy use in commercial buildings. After this he joined the Centre for Mathematical Analysis at the Australian National University as a research fellow and lecturer: there he taught and did theoretical and applied work on inverse problems in optics, atmospheric chemistry and groundwater modelling.

Since 1995 Dr Newsam has been a Principal Research Scientist with DSTO, leading a group now within Surveillance Systems Division tackling problems in image analysis and related areas. His current research interests include image processing and compression, spline fitting and image registration, reconstructing shapes from imagery, modelling and generating random fields, and Radon transforms.

Nicholas J. Redding
Surveillance Systems Division

Nicholas Redding received a B.E. and Ph.D. in electrical engineering all from the University of Queensland, Brisbane, in 1986 and 1991, respectively. From 1988 he received a Research Scientist Fellowship from the Australian Defence Science and Technology Organisation (DSTO) and then joined DSTO in Adelaide as a Research Scientist after completing his Ph.D. in artificial neural networks in 1991. In 1996 he was appointed as a Senior Research Scientist in the Microwave Radar Division (now Surveillance Systems Division) of DSTO. Since joining DSTO he has applied image processing techniques to the automatic classification of ionospheric data, and more recently has researched target detection (both human and algorithmic) in synthetic aperture radar imagery. He has recently returned from a one and a half year posting to the UK's Defence Evaluation and Research Agency where he continued the development of a suite of target detection algorithms for SAR imagery and researched new algorithms in SAR image forming using the circular Radon transform.

Contents

Figures

1 Introduction

Many problems in computer vision and elsewhere centre on finding some member of a family of curves that best fits a given set of points in a plane (or, more generally, on finding a member of an m-dimensional family of surfaces that best fits a set of points in n-dimensional space). At present the preferred method for this is *orthogonal distance regression*: in this approach the best fit is taken to be the curve which minimises the sum of squares of distances from the data points to their nearest neighbours on the curve. Orthogonal distance regression is well known to be computationally difficult, and although some good algorithms are now available [8, 1], it can still be very hard to numerically determine the optimal fit. This paper is not concerned with efficient computation, however. Rather it explores the more fundamental problem of the flaws in the stochastic model on which orthogonal distance regression is based, and develops an improved model and associated distance measures that produce better fits.

The paper stemmed from the authors' practical experience in fitting curves to ionogram data (see [14, 15, 16, 17] and Appendix A for a brief discussion of ionograms and their approximation), while the approach taken was inspired by some perceptive remarks by Kanatani on the distinction between geometric and statistical distance measures in fitting curves to data [11, 12, 10, p. 359]. Figure 1 shows a typical trace in an ionogram, together with an idealized fit to the data. It was found that curves fitted to these data sets by orthogonal distance regression had a tendency to underfit the nose of the trace (*i.e.*, not to extend as far as the data). Theoretical confirmation of this bias in the fitting procedure would have serious implications for ionogram modelling due to the importance of accurately locating the exact extent of the nose in applications of interest.

The tendency of orthogonal distance regression to underfit corners has been noted previously. Kanatani's intuitive explanation for this is that the sum of squared distances is primarily a geometric measure of fit and does not account for all the stochastic processes that play a part in generating real data. To see this, consider an simple example in which the data are noisy measurements of points that are sampled randomly along some underlying curve with a nose like that in figure 1. Then data points are more likely to be found inside the nose than outside it, since a point inside the nose at a fixed distance from the curve will be closer on average to the rest of the curve than is a point at the same distance but on the outside of the curve. This bias is a product of the measurement process, the distribution of samples on the true curve, and the curvature of the nose: the sum of squared distances does not take all of these issues fully into account. In particular, a curve chosen under this metric is likely to underfit the true nose since there are likely to be more data points inside the nose than outside.

Kanatani's insight prompted us to develop the simple but reasonable statistical model presented here for the distribution of a set of noisy observations as a function of the underlying curve. We then derive the likelihood function for observed data under this model: while it cannot be exactly evaluated for anything other than the simplest curves, it lends itself naturally to asymptotic approximations. In particular we show that under this model orthogonal distance regression can be viewed as a first order (*i.e.*, linear) approximation to the true maximum likelihood estimator. Moreover a second order approximation gives an estimator that is a simple extension of orthogonal regression that now includes con-

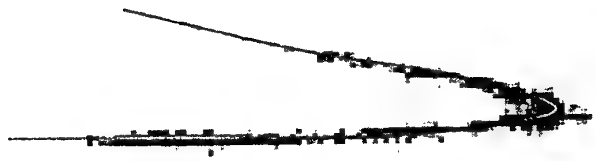


Figure 1: An ionogram trace overlayed with an idealized fit using the curve c from some family C of curves.

tributions from the curvatures at the closest points on the curve: these contributions compensate for the tendency already noted to underfit corners.

Kanatani's argument also suggests that orthogonal distance regression will fail to accurately reconstruct globally convex curves, although in this case the argument now implies that the regression will overfit the true curve. To see this, note that a random perturbation of a point on a convex curve is more likely to fall outside the curve than within it. Therefore, under the simple model proposed above, more data points are likely to lie outside the curve than inside, so orthogonal distance regression is likely to produce an estimate that lies outside the actual curve. While this behaviour does not seem to have been explicitly noted in the literature before, it is certainly apparent in the examples considered here. Once again, however, the problem is addressed by a proper statistical analysis. In the model presented here the likelihood function turns out to include a term involving the length of the curve being fitted. This term favours shorter curves, counterbalancing the asymmetry in the distribution of observed points.

More formally, it is known (see, *e.g.*, [6, p. 247]), if not widely appreciated, that orthogonal distance regression may be both *biased* and *inconsistent*. By this we mean that the average of fits derived from repeated finite samples of a given underlying curve may not be the underlying curve, and that as the number of samples tends to infinity while the noise level stays the same the fitted curve may not converge to the true curve. One explanation for this is that orthogonal distance regression is maximum likelihood estimation under a simple statistical model in which the location of the samples on the true curve are unknown parameters. Since the number of free parameters in this model grows in proportion to the number of data points, none of the classical convergence results for maximum likelihood estimation apply. Another way of viewing the model presented here is that it extends the simple model in a way that allows elimination of these free parameters

The structure of the paper is as follows. In the next section we introduce notation and describe the standard stochastic model of data formation: the observed data are noisy perturbations of unknown points on some underlying true curve. We then extend this model with the additional assumption that the unknown points are uniformly distributed along the true curve. Within each model we can then define the best fitting curve as that curve which maximises the likelihood of observing the given data. Under Gaussian noise, for the first model this maximisation is just orthogonal distance regression; in the second model the likelihood function for each observation takes the form of an integral of a Gaussian kernel along the given curve, with the maximum of the kernel being at the closest

point on the curve to the observation. While this integral is intractable in general, its form immediately suggests approximation by an asymptotic expansion about the closest point.

Section 3 derives the simplest such approximation by taking a locally linear approximation to the curve at the closest point and assuming that: the data points are independent samples; the variation in curve length is insignificant; and the radius of curvature is large compared to the noise level. Under this approximation the maximisation reduces to orthogonal distance regression. This shows maximum likelihood estimation under the new model is a credible extension of orthogonal distance regression.

Section 4 then derives an improved estimate based on a quadratic approximation of the curve at the closest point, and shows that this leads to the addition of a linear correction to each squared distance. This correction term is product of the distance to the curve and the ratio of the noise variance and the local curvature. Thus, unlike standard orthogonal distance regression, the penalty function requires knowledge of the variance: this is usually unknown, but can be estimated by including it as a variable in the likelihood maximisation and solving for it explicitly.

Section 5 elucidates the differences between the fits calculated from the various approximations derived in sections 3 & 4 for the special case of fitting a circle centred on the origin to observed data. In this case analytic expressions can be derived for the various maximum likelihood estimators, along with expressions for the bias and variance in these estimators. The bias is expressed as a series in the ratio of the variance to the square of the radius, and is independent of the number of observations. Under orthogonal distance regression the bias expansion starts with a first order term: in the new estimator this term is eliminated from the expansion. We also determine the conditions under which the bias in orthogonal distance regression is large compared to the variance, and so under which it would pay to adopt the more accurate estimator proposed here.

Section 6 gives some more realistic illustrative numerical results on fitting ellipses. These show the tendency of orthogonal distance regression to overfit convex curves, and that the second order estimator presented here produces a more accurate fit. Finally, after conclusions in section 7, some supporting material is presented in the appendices. This includes: a brief description of the autoscaling problem for ionograms; closed form expressions for various moment integrals used in the main body; and an extension of the analysis to the problem of fitting an d–1 dimensional surface to a set of points in $\mathbb{R}^d$.

2 The Problem

In the following subsections, we firstly examine the standard model and show how it reduces to orthogonal distance regression, and secondly we extend the model to address the failings of the standard model.

2.1 The Standard Model

Formally, the problem of interest is the following:

Problem: Given a set of noisy observations $\mathbf{x}_1, \ldots, \mathbf{x}_N \in \mathbb{R}^2$, find the best fitting curve c to the data from some family $\mathcal{C}$ of curves.

As noted by Kanatani, a necessary precondition for any well-defined fitting procedure is a model for data formation and the introduction of errors. The usual model is [4, 12]:

Standard stochastic model: Each $\mathbf{x}_n$ is an independent noisy observation of a point $\mathbf{y}_n$ on the curve that has been corrupted by additive isotropic Gaussian noise with variance σ^2.

Note that this model has two distinct components, each of which needs to be determined in the fitting process. The first and obvious component is the unknown curve c: this will usually be represented as some function $c(\boldsymbol{\theta})$ of a vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_K)$ of parameters. The second, less obvious, component is the nature of the stochastic processes determining the observed distributions. This too will normally be restricted to some known class of models parameterised by a vector $\boldsymbol{\sigma} = (\sigma_1, \ldots, \sigma_L)$. The fitting process must then determine the parameters $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$.

For simplicity, for now we assume the noise has zero mean and is identically and independently distributed: thus it can be characterised solely by its standard deviation σ. In many practical situations, however, the noise will be correlated; for example, trace extraction in ionogram autoscaling and edge extraction both produce connected segments of pixels which are then to be fitted by lines or curves. In this case neighbouring pixels on the segments must be adjacent, and so are strongly correlated: how this correlation should be incorporated into the model will be considered in a later paper.

Thus, given a curve c and noise level σ, the standard model defines the conditional probability $p(\mathbf{x}|\mathbf{y}; c, \sigma)$ that a point $\mathbf{x}$ is a noisy observation of a point $\mathbf{y}$ on c as

$$p(\mathbf{x}|\mathbf{y}; c, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2}$$

subject to the constraint $\mathbf{y} \in c$. This gives the following joint distribution for data and model:

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{y}_1, \ldots, \mathbf{y}_N; c, \sigma) = \prod_{i=1}^{N} p(\mathbf{x}_n|\mathbf{y}_n; c, \sigma).$$

Under this model the curve c, noise level σ and unknown true points $\mathbf{y}_1, \ldots, \mathbf{y}_N$ are all parameters describing a distribution that are to be inferred from the available data. Maximum likelihood estimation has them being chosen to minimise the log-likelihood function

$$l(\mathbf{x}_1, \ldots, \mathbf{x}_N : \mathbf{y}_1, \ldots, \mathbf{y}_N, c, \sigma) = -\sum_{n=1}^{N} \ln p(\mathbf{x}_n|\mathbf{y}_n; c, \sigma) \tag{1}$$

giving the problem

$$\min_{\mathbf{y}_1, \ldots, \mathbf{y}_N; c, \sigma} \quad \frac{1}{2\sigma^2} \sum_{n=1}^{N} \|\mathbf{x}_n - \mathbf{y}_n\|^2 + 2N \ln \sigma \tag{2}$$
$$\text{subject to} \quad \mathbf{y}_1, \ldots, \mathbf{y}_N \in c.$$

It is clear that for fixed c the optimal choice for $\mathbf{y}_n$ is $\mathbf{y}_n = \mathbf{x}_n^*$, where $\mathbf{x}_n^*$ is the closest point on c to $\mathbf{x}_n$. Moreover, the $\mathbf{y}_n$ are independent of σ, and, if σ is unknown, optimising (2) with respect to σ gives that its maximum likelihood estimate is

$$\sigma^2 = \frac{1}{2N}\sum_{n=1}^{N} d_n^2. \tag{3}$$

Here $d_n = \|\mathbf{x}_n - \mathbf{x}_n^*\|$ is the distance from $\mathbf{x}_n$ to c. Substituting this expression for σ back into (2) and exponentiating reduces it to standard orthogonal distance regression.

While this analysis provides a statistical foundation for orthogonal distance regression, it also highlights the method's drawbacks. In particular, it explicitly requires estimation of the unknown true sample points $\mathbf{y}_n$ as well as c and σ. In practice the points $\mathbf{y}_n$ are of little interest in themselves, and a model for the reduced conditional density $p(\mathbf{x}_1, \ldots, \mathbf{x}_N|c, \sigma)$ would be preferable to one for $p(\mathbf{x}_1, \ldots, \mathbf{x}_N|\mathbf{y}_1, \ldots, \mathbf{y}_N; c, \sigma)$. The $\mathbf{y}_n$ can only be eliminated within the standard model, however, by strong assumptions such as the noise level being small compared to the curvature (see, *e.g.*, [13, 12]). We therefore propose an alternative model.

2.2 The Extended Model

Ideally curve estimation would be based on a complete statistical analysis that started from a model for the full joint distribution $p(\mathbf{x}, \mathbf{y}; c, \sigma)$. This can be factored as

$$p(\mathbf{x}, \mathbf{y}; c, \sigma) = p(\mathbf{x}|\mathbf{y}; c, \sigma)p(\mathbf{y}; c, \sigma).$$

The standard model stops here, providing an expression for $p(\mathbf{x}|\mathbf{y}; c, \sigma)$ but making no assumptions about the form of $p(\mathbf{y}; c, \sigma)$. We now extend the model by noting that, under the reasonable assumption that the curve generation, curve sampling and noise processes are independent, $p(\mathbf{y}; c, \sigma)$ can be further factored as

$$p(\mathbf{y}; c, \sigma) = p(\mathbf{y}|c)p(c)p(\sigma).$$

We do not propose a complete model here, as the prior distributions $p(c)$ and $p(\sigma)$ are very problem dependent and usually impossible to specify in advance. We can, however, propose a reasonable universal model for the sampling process $p(\mathbf{y}|c)$ that generates the points observed on the curve, in particular that it is uniform in arclength. Thus

$$p(\mathbf{y}|c) = \frac{ds}{\int_c ds} = \frac{ds}{|c|}$$

where ds is the element of arc length and $|c|$ is the length of c.

With this model the unknown true sample points can now be removed from the analysis by noting that

$$\begin{aligned} p(\mathbf{x}|c, \sigma) &= \int p(\mathbf{x}|\mathbf{y}, c, \sigma)p(\mathbf{y}|c)d\mathbf{y} \\ &= \frac{1}{2\pi\sigma^2|c|}\int_c e^{-\|\mathbf{x}-\mathbf{y}(s)\|^2/2\sigma^2} ds \end{aligned} \tag{4}$$

where the points $\mathbf{y}(s)$ along the curve have been parameterized by arc length s. This gives an expression for the conditional probability

$$p(\mathbf{x}_1, \ldots, \mathbf{x}_N | c, \sigma) = \prod_{i=1}^{N} p(\mathbf{x}_n | c, \sigma) \tag{5}$$

as a product of integrals that no longer involve the $\mathbf{y}_n$.

We are now in a position to define the maximum likelihood estimate of the best fitting curve c to the data. As usual we define the log-likelihood function

$$l(\mathbf{x}_n : c, \sigma) \equiv -\ln p(\mathbf{x}_n | c, \sigma).$$

Therefore the maximum likelihood estimate of the true underlying curve c is the curve c^* that minimizes

$$l(\mathbf{x}_1, \ldots, \mathbf{x}_N : c, \sigma) = -\sum_{n=1}^{N} \ln p(\mathbf{x}_n | c, \sigma). \tag{6}$$

It now remains to evaluate the terms in (6). In practice it will be impossible to get closed form expressions for $l(\mathbf{x}_n : c, \sigma)$ for all but the simplest of curves, so we shall have to resort to asymptotic expansions of the integral in (4). But before starting this, we first wish to make a number of further comments on the model as proposed so far.

First, note that the above analysis shows that the a posteriori probability of a point $\mathbf{y}$ on a curve c generating a given observation $\mathbf{x}$ is

$$\begin{aligned} p(\mathbf{y} | \mathbf{x}, c, \sigma) &= \frac{p(\mathbf{x} | \mathbf{y}, \sigma) p(\mathbf{y} | c)}{p(\mathbf{x} | c, \sigma)} \\ &= \frac{e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2}}{\int_c e^{-\|\mathbf{x}-\mathbf{y}(s)\|^2/2\sigma^2} ds}. \end{aligned}$$

Second, in many cases the admissable curves c will have infinite length, for example when $\mathcal{C}$ is a family of straight lines. In this case we shall model $p(\mathbf{y}|c)$ by an *non-informative prior* [3]. In this model we assume all points on c are equally likely and all curves in $\mathcal{C}$ have the same length, and we interpret these statements to mean that the factor $|c|$ can simply be removed from (6).

Third, in principle the ideal quantity to compute is the a posteriori density

$$p(c, \sigma | \mathbf{x}_1, \ldots, \mathbf{x}_N) = p(\mathbf{x}_1, \ldots, \mathbf{x}_N | c, \sigma) \frac{p(c, \sigma)}{p(\mathbf{x}_1, \ldots, \mathbf{x}_N)}.$$

Knowledge of $p(c, \sigma | \mathbf{x}_1, \ldots, \mathbf{x}_N)$ would give a much better overall understanding of the problem, and would allow construction of Bayesian estimators that minimised appropriate loss functions. Note that in practice the curve generation and noise generation processes will most likely be independent, so that $p(c, \sigma) = p(c)p(\sigma)$.

Unfortunately computation of $p(c, \sigma | \mathbf{x}_1, \ldots, \mathbf{x}_N)$ requires knowledge of the prior distributions $p(\mathbf{x}_1, \ldots, \mathbf{x}_N)$, $p(c)$ and $p(\sigma)$, as well as $p(\mathbf{x}_1, \ldots, \mathbf{x}_N | c, \sigma)$. In some cases this

may be available. For example, often there will be no preferred origin in the problem, so that the family $\mathcal{C}$ includes all translates of itself. In this case $p(\mathbf{x}_1, \dots, \mathbf{x}_N)$ will be uniform regardless of $p(c)$ and $p(\sigma)$. On the other hand, if, as is often the case, the observations are first scaled and translated to the unit square before fitting, then $p(\mathbf{x}_1, \dots, \mathbf{x}_N)$ will not be uniform but will reflect the prior distribution of curves $p(c)$.

More importantly, it is not at all clear what form a reasonable prior distribution $p(c)$ on the space of all curves might take, or how to define it. Possibilities exist, such as Weiner measures on spaces of curves, but a discussion of them is well beyond the scope of this paper.

Finally, as noted before, in practice in most fitting problems the class of curves $\mathcal{C}$ consists of a family of shapes $c(\boldsymbol{\theta})$ parameterized by the vector $\boldsymbol{\theta} = (\theta_1, \dots, \theta_K)$. In this case maximum likelihood estimation now becomes maximisation of $p(\mathbf{x}_1, \dots, \mathbf{x}_N|\boldsymbol{\theta})$ over $\boldsymbol{\theta}$, while the expression for the posterior distribution is

$$p(\boldsymbol{\theta}|\mathbf{x}_1, \dots, \mathbf{x}_N) = p(\mathbf{x}_1, \dots, \mathbf{x}_N|\boldsymbol{\theta}) \frac{p(\boldsymbol{\theta})}{p(\mathbf{x}_1, \dots, \mathbf{x}_N)}.$$

Knowledge of the posterior now requires knowledge of the prior distribution on the parameter space. This can be even more problematical than knowledge of the prior on some space of curves: with a poor (or even a natural) parameterization, the natural measure on the parameter space may not be very closely related to the natural measure on the space of actual curves. Nevertheless the posterior distribution $p(\boldsymbol{\theta}|\mathbf{x})$ would still be the preferred distribution to work with if any further analysis is to be carried out, such as hypothesis testing to decide if curves can be merged.

3 First order approximation

To compute the likelihood function we need to evaluate integrals of the form

$$p(\mathbf{x}_n|c, \sigma) = \frac{1}{2\pi\sigma^2 |c|} \int_c e^{-\|\mathbf{x}_n - \mathbf{y}(s)\|^2/2\sigma^2} ds. \tag{7}$$

In practice this will be impossible to calculate exactly for all but the simplest of curves. The Gaussian kernel, however, naturally suggests some form of asymptotic approximation along the general lines of Laplace's method [5] around the maximum of the integrand. This occurs at the point

$$\mathbf{x}_n^* \equiv \arg\min\{\|\mathbf{x}_n - \mathbf{y}(s)\|^2 : \mathbf{y}(s) \in c\}$$

i.e., $\mathbf{x}_n^*$ is the closest point to $\mathbf{x}_n$ on c. We shall denote the distance $\|\mathbf{x}_n - \mathbf{x}_n^*\|$ by d_n.

We begin by constructing a simple linear approximation to the integral in (7): we shall show that under this maximum likelihood estimation again reduces to orthogonal distance regression. Let $\mathbf{n}_n$ be the normal to c at $\mathbf{x}_n^*$, and let $T_n \equiv \{\mathbf{y} : \mathbf{n}_n^{\mathrm{T}}(\mathbf{y} - \mathbf{x}_n^*) = 0\}$ be the tangent line to c. Thus, if the curve is smooth and varies very little over the central support of the Gaussian kernel (*i.e.*, the radius of curvature is large compared to σ), then we may approximate the integral in (7) by

$$p(\mathbf{x}_n|c, \sigma) \sim \frac{1}{2\pi\sigma^2} \frac{1}{|c|} \int_{\mathbf{y} \in T_n} e^{-\|\mathbf{x}_n - \mathbf{y}\|^2/2\sigma^2} ds.$$

Since $\mathbf{x}_n^*$ is the closest point to $\mathbf{x}_n$ on c, $\mathbf{x}_n - \mathbf{x}_n^*$ must be perpendicular to T_n, *i.e.*, $\mathbf{x}_n - \mathbf{x}_n^*$ must be proportional to $\mathbf{n}_n$. Therefore by Pythagoras' theorem for $\mathbf{y} \in T_n$,

$$\|\mathbf{x}_n - \mathbf{y}\|^2 = \|\mathbf{x}_n - \mathbf{x}_n^*\|^2 + \|\mathbf{x}_n^* - \mathbf{y}\|^2,$$

from which it follows that

$$\begin{aligned} p(\mathbf{x}_n|c,\sigma) &\sim \frac{1}{2\pi\sigma^2}\frac{e^{-d_n^2/2\sigma^2}}{|c|}\int_{\mathbf{y}\in T_n} e^{-\|\mathbf{x}_n^*-\mathbf{y}\|^2/2\sigma^2}ds \\ &= \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-d_n^2/2\sigma^2}}{|c|}. \end{aligned}$$

Thus the log-likelihood function reduces to

$$l(\mathbf{x}_1,\dots,\mathbf{x}_N : c,\sigma) \sim \sum_{n=1}^{N}\frac{d_n^2}{2\sigma^2} + N\left[\ln|c| + \ln\sqrt{2\pi}\sigma\right]. \tag{8}$$

This loss function is identical with that used in orthogonal distance regression, except for the additional contributions due to the variable curve length and the unknown noise level. If both of these were known and fixed, then the linear approximation to maximum likelihood estimation would indeed reduce to orthogonal distance regression. In practice, however, both terms will vary, and, as shall be seen in section 5, their inclusion can make a significant difference to the results produced by straight orthogonal distance regression.

In further analysing the contributions from the terms involving the variance and curve length to the maximum likelihood estimator, note first that the variance must be known or estimated if the curve length contribution is included. This follows since the relative weight of the contributions from the sum of squared differences and log of curve length is moderated by the variance. Fortunately, if the variance is unknown it can be eliminated by direct substitution. To see this, note that at a minimum of l, $\partial l/\partial\sigma = 0$ must hold regardless of the form of c. This implies

$$\sigma^2 = \frac{1}{N}\sum_{n=1}^{N} d_n^2.$$

Substituting this back into (8), disregarding constants, exponentiating and then squaring shows that the first order approximate maximum likelihood estimate of c is the minimiser of

$$e_1(\mathbf{x}_1,\dots,\mathbf{x}_N : c) \equiv |c|^2\sum_{n=1}^{N} d_n^2. \tag{9}$$

While this minimisation is still very close to orthogonal distance regression, curve length now obviously plays a significant role.

The effect of the curve length contribution has a reasonable interpretation: all other things being equal (*i.e.*, if the closest point remains unchanged and the first order asymptotic approximation is still valid), a shorter curve has a smaller range of $\mathbf{y}$ values that could have given rise to the observation and therefore an increased probability that a

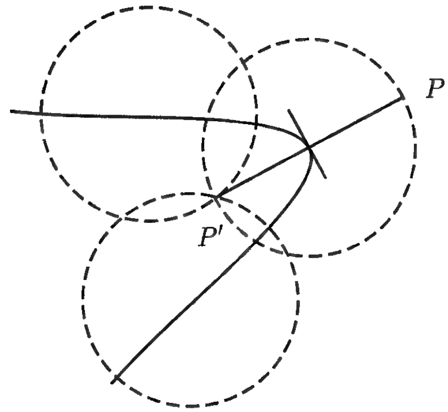


Figure 2: Following [10], it can be seen that points on the curve with added independently and identically distributed noise are more likely to fall in the concave region containing P' than in the convex region containing P. The three circles shown with centre on the curve represent contours of constant probability under the standard stochastic. Obviously, more points on the curve are likely to contribute to noisy observations at P' than at P. Consequently, an observation is more likely to occur at P' and similarly for other points in a concave region when compared with those in a convex region.

neighbourhood of the closest point was indeed the source of the observation. And while $e_1(\mathbf{x}_1, \ldots, \mathbf{x}_N : c)$ can be decreased to zero if $|c| \downarrow 0$, this behaviour violates the assumptions under which (8) and (9) were derived, as the curvature will no longer be large with respect to the noise level at any point $\mathbf{x}_n^*$.

Strictly speaking, a full one-dimensional approximation would see the terms involving $|c|$ disappear from (8) and (9): since tangent lines have infinite length and the length of all tangents is the same, as noted in the previous section it would be reasonable to use a non-informative prior. This has its attractions, especially when $|c|$ is difficult to compute and the curves under consideration have infinite length and no well defined boundary. It conflicts with strict maximum likelihood esimation, however, and, as section 5 shows, terms involving $|c|$ can make significant contributions.

Finally we note again that in practice $l(\mathbf{x}_1, \ldots, \mathbf{x}_N : c, \sigma)$ will usually be represented in parameterised form as $l(\mathbf{x}_1, \ldots, \mathbf{x}_N : \boldsymbol{\theta}, \sigma)$, and the problem actually faced will be to minimise (8) or (9) as functions of $\boldsymbol{\theta}$.

4 Second Order Approximation

The problem with the first order model in that it does not account for the effect that local curvature is likely to have on the distribution of observed points. For example, consider the following situation of two points P and P' that are equidistant to and on either side of the nose of a curve, and such that the nose is the closest point on the curve to both points (figure 2, [10]). The positive curvature means that point P' is more likely to be generated by the noise process than point P since P' is closer to the legs of the curve. Thus while each point will make an equal but opposite contribution to locating the

nose in orthogonal distance regression, the point P' is more likely to be observed so the nose is more likely to contract towards P' than P. Under orthogonal distance regression both points will have an equal but opposite effect in positioning the curve, but the unequal probability of their occurence means that the fitted curve is more likely to lie inside the true nose than outside. This bias is a product of the prior distribution of observations and the curvature of the nose: it is not properly accounted for in minimising the sum of squared distances.

A similar argument suggests that orthogonal distance regression also gives a biased result when fitting a closed convex curve to noisy data. To see this, note that a random perturbation of a point on such a curve is more likely to fall outside the curve than within it. Moreover points falling outside are more likely to be further away from the true curve than points falling within it. Therefore orthogonal distance regression on the observed data is likely to return a biased fit that lies outside the actual curve. We shall show later that this is indeed the case.

The above arguments, that orthogonal distance regression based on the linear approximation may produce biased fits, are supported by our observation that curves derived using orthogonal distance regression can be biased around the nose of traces when fitting ionograms [14, 15, 16, 17]. Therefore we seek a better approximation to the log-likelihood function that will take into account the effect of local curvature.

4.1 Deriving the second order approximation

Let $c_n^* \equiv c_n^*(r_n, \mathbf{z}_n)$ be the best approximating circle to c at $\mathbf{x}_n^*$, and suppose that c_n^* has centre $\mathbf{z}_n$ and radius r_n (r_n is the inverse of the curvature of c at $\mathbf{x}_n^*$). Again $\mathbf{x}_n - \mathbf{x}_n^*$ is perpendicular to the tangent of c at $\mathbf{x}_n^*$; as the tangent is also perpendicular to the radius of curvature, $\mathbf{x}_n - \mathbf{x}_n^*$ will be proportional to $\mathbf{x}_n^* - \mathbf{z}_n$. Without loss of generality, we may translate and rotate the plane so that the global coordinate system coincides with a local coordinate system in which

$$\begin{aligned}
\mathbf{z}_n &= (0,0), \\
\mathbf{x}_n &= (r_n - d_n, 0), \\
\mathbf{x}_n^* &= (r_n, 0).
\end{aligned}$$

We shall adopt the convention that d_n is signed. This will indicate whether c is concave or convex with repect to $\mathbf{x}_n$: if d_n is positive the $\mathbf{x}_n$ lies "inside" (within the locally convex region defined by) c, otherwise $\mathbf{x}_n$ lies "outside" the curve. (It is perhaps more natural to take r_n to be signed and to use this sign to indicate the local geometry, and indeed this convention is adopted in the extension of these results to $\mathbb{R}^d$ presented in appendix C. Nevertheless we prefer to take d_n to be signed here as it simplifies the form of certain integrals that appear below.)

With this approximation we have

$$\begin{aligned}
p(\mathbf{x}_n|c,\sigma) &\sim \frac{1}{2\pi\sigma^2}\frac{1}{|c|}\int_{c_n^*} e^{-\|\mathbf{x}_n - \mathbf{y}(s)\|^2/2\sigma^2} ds \\
&= \frac{1}{2\pi\sigma^2}\frac{1}{|c|}\int_0^{2\pi} e^{-\|\mathbf{x}_n - (\mathbf{z}_n + r_n(\cos\theta,\sin\theta))\|^2} r_n d\theta
\end{aligned}$$

and

$$\begin{aligned}\|\mathbf{x}_n - (\mathbf{z}_n + r_n(\cos\theta, \sin\theta))\|^2 &= \|(r_n - d_n, 0) - (r_n\cos\theta, r_n\sin\theta)\|^2 \\ &= (r_n - d_n - r_n\cos\theta)^2 + r_n^2\sin^2\theta \\ &= 2r_n(r_n - d_n)(1-\cos\theta) + d_n^2.\end{aligned}$$

The integral can be be evaluated exactly as a Bessel function, giving

$$\begin{aligned}p(\mathbf{x}_n|c,\sigma) &\sim \frac{1}{2\pi\sigma^2}\frac{r_n}{|c|}e^{-d_n^2/2\sigma^2}e^{-r_n(r_n-d_n)/\sigma^2}\int_0^{2\pi} e^{\frac{r_n(r_n-d_n)\cos\theta}{\sigma^2}}\,d\theta \\ &= \frac{1}{\sigma^2}\frac{r_n}{|c|}e^{-d_n^2/2\sigma^2}e^{-r_n(r_n-d_n)/\sigma^2} I_0\left(\frac{r_n(r_n-d_n)}{\sigma^2}\right).\end{aligned} \tag{10}$$

We now look at an asymptotic expansion of $p(\mathbf{x}_n|c,\sigma)$ under the assumptions

$$\frac{r_n}{\sigma} \gg 1, \qquad \frac{r_n}{d_n} \gg 1, \qquad \text{and} \qquad \frac{d_n}{\sigma} = \mu \sim 1.$$

These assumptions are consistent with the underlying model: we expect that the distance between a typical observation $\mathbf{x}_n$ and the closest point on the curve to it will be of order σ. Thus, after defining $\rho \equiv r_n/\sigma$, we seek an asymptotic expansion of

$$\phi(\rho) \equiv \rho e^{-\rho(\rho-\mu)} I_0(\rho(\rho-\mu))$$

for large ρ. To order ρ^{-2} we have

$$\phi(\rho) \sim \rho e^{-\rho(\rho-\mu)} \frac{e^{\rho(\rho-\mu)}}{\sqrt{2\pi\rho(\rho-\mu)}}\left[1 + \frac{1}{8}\frac{1}{\rho(\rho-\mu)}\right],$$

giving

$$\begin{aligned}-\ln\phi(\rho) &\sim \ln\sqrt{2\pi} + \frac{1}{2}\ln\left[1 - \frac{\mu}{\rho}\right] - \ln\left[1 + \frac{1}{8\rho^2}\frac{1}{(1-\mu/\rho)}\right] \\ &\sim \ln\sqrt{2\pi} - \frac{1}{2}\left[\frac{\mu}{\rho} + \frac{1}{2}\left(\frac{\mu}{\rho}\right)^2\right] - \frac{1}{8\rho^2},\end{aligned} \tag{11}$$

and therefore that

$$-\ln p(\mathbf{x}_n|c,\sigma) \sim \ln\sqrt{2\pi}\sigma + \ln|c| + \frac{d_n^2}{2\sigma^2} - \frac{d_n}{2r_n} - \frac{1}{8r_n^2}[\sigma^2 + 2d_n^2]. \tag{12}$$

It is not obvious that the coefficients of the second order terms σ^2/r_n^2 and d_n^2/r_n^2 in the above equation are correct as is, or whether they would be altered by a more accurate approximation to c that involved higher derivatives. Therefore we shall only include the first order term to give

$$l(\mathbf{x}_1, \dots, \mathbf{x}_N : c, \sigma) \sim \sum_{n=1}^{N}\left[\frac{d_n^2}{2\sigma^2} - \frac{d_n}{2r_n}\right] + N\left[\ln|c| + \ln\sqrt{2\pi}\sigma\right]. \tag{13}$$

(However see the end of the next subsection for a discussion of an alternative representation of the log-likelihood function that includes some second order terms).

Equation (13) has one significant difference from (8): the linear correction term $d_n/2r_n$. This will act to counter the problem of underfitting noses noted earlier. For if $\mathbf{x}_n$ lies inside the curve then d_n is positive, so the decrease in the term d_n/r_n will offset to some degree the increase in $d_n^2/2\sigma^2$ as the curve is moved away from $\mathbf{x}_n$. In contrast, if $\mathbf{x}_n$ lies outside the curve then both $d_n^2/2\sigma^2$ and d_n/r_n will be positive, and so will attract the curve towards $\mathbf{x}_n$.

4.2 Alternative forms for the second order approximation

We now present some alternative formulations of (13) that may be more appropriate for computational purposes, or that give further insight into the underlying concepts. First, an unknown variance can again be eliminated from (13) by direct substitution of the maximum likelihood estimate of c. Since σ does not appear in the linear term, the condition $\partial l/\partial\sigma = 0$ implies the maximum likelihood estimate of the variance is $\sigma^2 = \frac{1}{N}\sum_{n=1}^{N} d_n^2$. Substituting this back into (13), disregarding constants, exponentiating and then squaring gives that the second order approximate maximum likelihood estimate of c is the minimiser of

$$e_2(\mathbf{x}_1,\ldots,\mathbf{x}_N : c) \equiv |c|^2 \cdot \left(\frac{1}{N}\sum_{n=1}^{N} d_n^2\right) \cdot \exp\left[-\frac{1}{N}\sum_{n=1}^{N}\frac{d_n}{r_n}\right]. \tag{14}$$

Next, it is worth noting that slightly different versions of (13) and (14) can be derived by going back to (11) and retaining the term $\ln(1-\mu/\rho)$ as is, rather than replacing it by a first order approximation. This gives

$$l(\mathbf{x}_1,\ldots,\mathbf{x}_N : c,\sigma) \sim \sum_{n=1}^{N}\left[\frac{d_n^2}{2\sigma^2} + \frac{1}{2}\ln\left(1-\frac{d_n}{r_n}\right)\right] + N\left[\ln|c| + \ln\sqrt{2\pi}\sigma\right],$$

and

$$e_2(\mathbf{x}_1,\ldots,\mathbf{x}_N : c) = |c|^2 \cdot \left(\frac{1}{N}\sum_{n=1}^{N} d_n^2\right) \cdot \left[\prod_{n=1}^{N}\left(1-\frac{d_n}{r_n}\right)\right]^{1/N}.$$

Nevertheless, while these equations may be more accurate approximations to the true log-likelihood and associated functions in principle, in practice (13) and (14) are equivalent to first order and are also easier to minimise numerically since they allow d_n/r_n to be greater than 1 without giving rise to singularities or similar difficulties.

Finally it is instructive to explore the form of (13) a little further. In particular, it is tempting to complete the square and rewrite the equation as

$$l(\mathbf{x}_1,\ldots,\mathbf{x}_N : c,\sigma) \sim \sum_{n=1}^{N}\frac{(d_n-\sigma^2/2r_n)^2}{2\sigma^2} + N\left[\ln|c| + \ln\sqrt{2\pi}\sigma\right]. \tag{15}$$

This can then be interpreted as the log-likelihood function of a normal distribution approximation to $p(\mathbf{x}_n|c,\sigma)$. This interpretation, however, would be wrong and misleading, as well as not agreeing to second order with (12). To understand why it suffices to note that, to first order, $E[d_n] \sim -\sigma^2/2r_n$: this follows from setting $\rho_n = r_n - d_n$ and appealing to (B2) of Appendix B. (Here $E[x]$ denotes, as usual the expectation of the random variable x.) Thus the term $\sigma^2/2r_n$ in (15) cannot be interpreted as the mean of $p(\mathbf{x}_n|c,\sigma)$ as it has the wrong sign. Rather $p(\mathbf{x}_n|c,\sigma)$ is a skewed distribution and $\sigma^2/2r_n$ is a first order approximation to its maximum (or mode): as the mode is the relevant statistic in maximum likelihood estimation, the approximation has instead constructed a local quadratic approximation to the likelihood function about the mode rather than the mean. This confirms and clarifies both the observations made in the introduction: that, while on average an observation is more likely to be made outside the curve than inside, the most likely single point for any observation is just inside the curve.

5 Comparison of approximate maximum likelihood estimates of circles

To get a feel for the differences between the solutions of each of the various approximate maximum likelihood estimators introduced above, we apply them here to the particular problem of fitting a circle to noisy data. Let c_0 be a circle of radius r_0 centred on the origin, and let $\{\mathbf{x}_n\}_{n=1}^N$ be a collection of noisy observations of c_0 generated according to the basic model under noise level σ. We seek a circle c of radius r also centred on the origin that best fits this data: *i.e.*, we seek an estimate r^* of r_0.

To achieve this, we first represent the data in polar coordinates as $\mathbf{x}_n = (\rho_n \cos\theta_n, \rho_n \sin\theta_n)$. Then substituting $r_0 - \rho_n$ for d_n in (10) shows that the distribution of the $\mathbf{x}_n$ can be summarised in the density function

$$p(\mathbf{x}_n|c_0,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(r_0-\rho_n)^2/2\sigma^2} e^{-r_0\rho_n/\sigma^2} I_0\left(\frac{r_0\rho_n}{\sigma^2}\right) \rho_n \, d\rho_n \, d\theta_n. \tag{16}$$

We now use this distribution to calculate the estimates that will be returned by each of the approximations considered so far.

5.1 Orthogonal distance regression

The squared distance of $\mathbf{x}_n$ to any circle c of radius r centred on the origin is simply $(r - \rho_n)^2$. Thus the estimate of r_0 returned by orthogonal distance regression will be the solution r^* of

$$\min_r \quad \sum_{n=1}^N (r - \rho_n)^2.$$

Thus

$$r^* = \frac{1}{N} \sum_{n=1}^N \rho_n. \tag{17}$$

Since the ρ_n are random variables, so also is r^*. In particular,

$$\begin{aligned} E[r^*] = E[\rho_n] &= \frac{1}{\sigma^2}\int_0^\infty \rho e^{-(\rho-r_0)^2/2\sigma^2} e^{-r_0\rho_n/\sigma^2} I_0\left(\frac{r_0\rho}{\sigma^2}\right)\rho\, d\rho \\ &= \frac{e^{-r_0^2/2\sigma^2}}{\sigma^2}\int_0^\infty \rho^2 e^{-\rho^2/2\sigma^2} I_0\left(\frac{r_0\rho}{\sigma^2}\right) d\rho. \end{aligned}$$

This integral is evaluated in Appendix B: under the assumption that $r_0/\sigma \gg 1$ (B2) gives

$$E[r^*] \sim r_0\left[1+\frac{1}{2}\left(\frac{\sigma}{r_0}\right)^2+\frac{1}{8}\left(\frac{\sigma}{r_0}\right)^4\right]. \tag{18}$$

Equation (18) shows that straight orthogonal distance regression will overestimate the true radius. This is not unexpected: as noted previously, since the circle is a convex curve a random perturbation of any point on it is more likely to fall outside it than inside.

In a similar vein the variance of the orthogonal distance regression estimate is

$$\mathrm{var}[r^*] = \frac{\mathrm{var}[\rho_n]}{N} = \frac{E[\rho_n^2]-E[\rho_n]^2}{N}.$$

Equation (B3) of Appendix B gives $E[\rho_n^2] = r_0^2 + 2\sigma^2$. Therefore retaining all terms to order $(\sigma/r_0)^4$ in (B2) gives

$$\begin{aligned} \mathrm{var}[r^*] &\sim \frac{1}{N}\left[r_0^2+2\sigma^2-r_0^2\left[1+\frac{1}{2}\left(\frac{\sigma}{r_0}\right)^2+\frac{1}{8}\left(\frac{\sigma}{r_0}\right)^4\right]^2\right] \\ &\sim \frac{\sigma^2}{N}\left(1-\frac{1}{2}\left(\frac{\sigma}{r_0}\right)^2\right) \end{aligned} \tag{19}$$

The variance estimate shows that if the bias in the orthogonal distance regression estimate for r_0 predicted by (18) is to be less than the estimate's variance (and thus not be the dominant term in the error in the estimate), then to first order N must satisfy

$$\frac{\alpha_p\sigma}{\sqrt{N}} \geqslant \frac{\sigma^2}{2r_0}$$

so that

$$\left(\frac{2\alpha_p r_0}{\sigma}\right)^2 \geqslant N. \tag{20}$$

Here α_p is the confidence level that ensures that the estimate r^* lies in the interval $[r_0 + \sigma^2/(2r_0) - \alpha_p\sigma/\sqrt{N}, r_0 + \sigma^2/(2r_0) - \alpha_p\sigma/\sqrt{N}]$ with probability p.

5.2 First order estimate with known variance

We now present the estimates r^* that will be returned by the various approximations to the maximum likelihood estimator derived in previous sections. Given a finite number N of observations, all of these estimates will be random variables. They are, however, now solutions of quadratics or higher order equations, and so have more complicated forms than (17). Therefore we shall not attempt to determine their exact means and variances here. Instead we shall simply determine the asymptotic values of the estimators as $N \uparrow \infty$ and note that the associated variances will again asymptote to σ^2/N. This analysis will be sufficient to illustrate the differences between the various estimators.

We start by minimising (8) under the assumption the noise level σ is known. Dividing by N, taking the limit as $N \uparrow \infty$ in (8) and disregarding constants gives a corresponding estimate of r_0 as the solution r^* of

$$\min_r \quad \frac{r^2 - 2rE[\rho_n] + E[\rho_n^2]}{2\sigma^2} + \ln r.$$

It is straightforward to show that

$$r^* = \frac{E[\rho_n] + (E[\rho_n]^2 - 4\sigma^2)^{1/2}}{2}.$$

Making the usual assumption that $r_0/\sigma \gg 1$ and substituting in the asymptotic expansion for $E[\rho_n]$ given in (18) shows that to fourth order

$$r^* \sim r_0 \left[1 - \frac{1}{2}\left(\frac{\sigma}{r_0}\right)^2 - \frac{3}{8}\left(\frac{\sigma}{r_0}\right)^4\right]. \tag{21}$$

Thus, if the variance is known and the proposed stochastic process for generating observations is correct, including a contribution from curve length in orthogonal distance regression will, to second order, see the resulting estimate undershoot the true radius by as much as the original orthogonal distance regression estimate overshot the mark.

5.3 First order estimate with unknown variance

If the noise variance is unknown then (9) can be used. For circle fitting, taking the limit as $N \uparrow \infty$ in (9) and disregarding constants gives

$$\min_r \quad r^2(r^2 - 2rE[\rho_n] + E[\rho_n^2]).$$

Again it is straightforward to show that

$$r^* = \frac{3E[\rho_n] + (9E[\rho_n]^2 - 8E[\rho_n^2])^{1/2}}{2}.$$

Under the assumption that $\sigma/r_0 \ll 1$, using (B3) and (18) gives that to fourth order

$$r^* \sim r_0 \left[1 - \frac{1}{2}\left(\frac{\sigma}{r_0}\right)^2 - \frac{7}{8}\left(\frac{\sigma}{r_0}\right)^4\right]. \tag{22}$$

Thus, if the variance is unknown, substituting the maximum likelihood variance estimate gives an estimate that agrees to first order with the estimate that uses the known variance.

5.4 Second order estimates with known variance

We now derive the estimate of r^* given by a second order approximation when the variance is known. Dividing by N in (13), then taking the limit as $N \uparrow \infty$ and disregarding constants gives a corresponding estimate of r_0 as the solution r^* of

$$\min_r \quad \frac{r^2 - 2rE[\rho_n] + E[\rho_n^2]}{2\sigma^2} - \frac{r - E[\rho_n]}{2r} + \ln r.$$

Thus r^* must be a root of

$$r^3 - r^2E[\rho_n] + r\sigma^2 - \frac{\sigma^2 E[\rho_n]}{2} = 0. \tag{23}$$

We next make the same assumptions and substitutions for $E[\rho_n]$ as in the corresponding first order case. This is followed by expanding r as a series in powers of $(\sigma/r_0)^2$ with unknown coefficients, substituting this series into (23) and then choosing the unknown coefficients to ensure that the first few terms in the result are identically zero. This gives that to fourth order

$$r^* \sim r_0 \left[1 + \frac{3}{8}\left(\frac{\sigma}{r_0}\right)^4\right]. \tag{24}$$

Thus, as expected, using the second order approximation to the log-likelihood function eliminates the leading order term in the bias of the estimate of the true radius r_0.

5.5 Second order estimate with unknown variance

If the noise variance is unknown then (14) can be used. For circle fitting, taking the limit as $N \uparrow \infty$ in (14) and disregarding constants gives the estimate of r_0 is the solution r^* of

$$\min_r \quad r^2(r^2 - 2rE[\rho_n] + E[\rho_n^2])e^{E[\rho_n]/r}.$$

Therefore r^* must be a root of

$$4r^3 - 6r^2E[\rho_n] + 2E[\rho_n^2] - E[\rho_n](r^2 - 2rE[\rho_n] + E[\rho_n^2]) = 0.$$

We again make the same assumptions and substitutions as in the first order estimate with unknown variance and the second order estimate above. This gives that to fourth order

$$r^* \sim r_0 \left[1 + \frac{1}{2}\left(\frac{\sigma}{r_0}\right)^4\right]. \tag{25}$$

Thus, even if the noise is unknown, using the maximum likelihood estimator of the variance still eliminates the leading order term in the bias.

5.6 Conclusions on circle fitting

We pause here to summarise the results presented in this section. The analysis clearly shows that, under a fixed noise level, orthogonal distance regression gives a biased estimate of the true radius. Moreover this estimate is inconsistent, in that the bias does not go to zero even if the number of data points $N \uparrow \infty$. Moreover the nature of the bias is such that orthogonal distance regression will tend to overfit the curve: *i.e.,* it will return an estimate r^* of the true radius r_0 with $r^* > r_0$. This bias is likely to be dominated, however, by the variance inherent in the fitting of noisy data unless N exceeds the inverse of the relative bias.

Including curve length in orthogonal distance regression over-compensates for the tendency to overfit. The leading order term in the bias expansion now has the same magnitude but opposite sign as the corresponding term in the bias expansion for the straight orthogonal distance regression estimate. Although the noise level σ must be known or estimated in this improved maximum likelihood estimation, the result is not greatly influenced by the use of the exact or an approximate value: both give the same leading order term in the expansion of the bias.

The leading order term can be removed, however, by using a second order approximation to the log-likelihood function. Again, while this approximation requires knowledge of the noise level, using the true value or an estimate makes little difference to the resulting estimate of the radius.

It is important to include the term involving curve length in the second order estimate, however: some simple calculations show that if this is dropped the resulting estimate is then even worse than that produced by straight orthogonal distance regression. This is not unexpected: as noted in section 4 the linear correction term tends to favour points lying on the outside of a convex curve. Given that the stochastic process underlying the formation of observations will produce more points outside the circle than inside, in the absence of a term favouring shorter curves the linear correction term will simply reinforce the tendency of straight orthogonal distance regression to overestimate the true radius.

Finally, the results in this section show that the tendency of orthogonal distance regression to underfit noses is not due to the convexity of the nose *per se*. Rather the change in curvature round the nose is the key factor that ensures that more points are likely to be observed inside the nose than outside. In a nose of constant curvature (*i.e.,* a circle), observations are more likely to fall outside the "corner" than inside.

6 Examples

In this section, we present the results of a number of simulations of fitting ellipses to noisy data. The data points were generated by adding identically distributed Gaussian noise to the x and y components of points that were uniformly sampled along the arc length of a given ellipse. To keep matters simple, the ellipse was centered on the origin and aligned with the x-y axes, so that the curve c was given by

$$c = \{f(x, y; a, b) = 1 \quad : \quad f(x, y; a, b) = \frac{x^2}{a^2} + \frac{y^2}{b^2}\}.$$

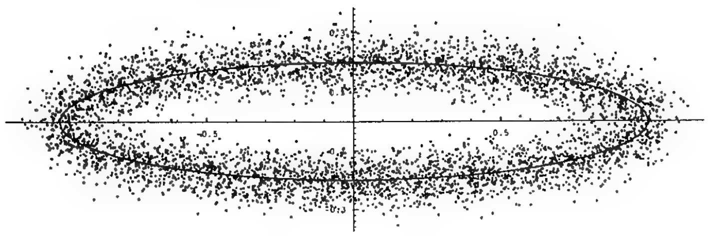
The noisy points from the ellipse were fitted by both orthogonal distance regression and minimising the log-likelihood function (13) under the assumptions that the variance is known. In each of the two test cases presented here, 5000 points were sampled uniformly along the arc length of a specified ellipse. Identically and independently distributed zero-mean Gaussian noise with $\sigma = 0.05$ was added to each of the x and y coordinates of the sampled points. In the first test case, the parameters of the ellipse were $a = 1$ and $b = 0.2$ (figure 3(a)), and in the more difficult second test case they were $a = 1$ and $b = 0.1$ (figure 4(a)).

The distances of the noisy points from the curve required to compute the sum of squared orthogonal distances (2) and the log-likelihood function (13) for particular ellipse parameter values a and b was found by computing the closest point on the ellipse using one of the techniques in [17]. The arc length and radius of curvature of a point on an ellipse is well known and can be obtained from many standard reference texts.

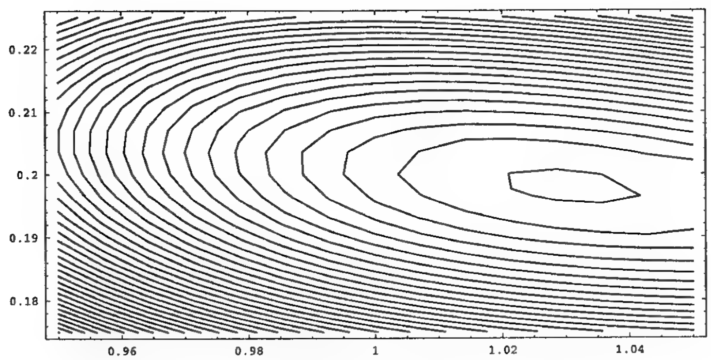
The contours of the sum of squared orthogonal distances and second-order approximation to the log-likelihood function as they vary with a and b are shown in figures 3(b) and 3(c) respectively for the first example, and figures 4(b) and 4(c) respectively for the second example. The bias in the orthogonal distance regression estimates is clearly evident. In fact, in the more difficult case (figure 4(b)), the minimum of the orthogonal distance regression metric does not have a well defined minimum. In contrast, the minimum of the second-order approximation to the log-likelihood function (13) is better defined and has a bias that is an order of magnitude smaller, especially in the more difficult example of $a = 1$ and $b = 0.1$.

As argued earlier, orthogonal distance regression should overfit convex curves, and indeed the bias exhibited is towards uniformly larger ellipses. Unfortunately this bias is most pronounced at the ends with maximum curvature, contradicting the previous argument for underfitting of noses. This shows the strengths and weaknesses of intuitive arguments. While both are plausible they actually argue for effects in opposing directions, and in the present case it seems likely the bias is actually due to other aspects of the ellipses' geometry. Nevertheless bias clearly exists; equally clearly it is greatly reduced under the analysis and algorithm proposed here.

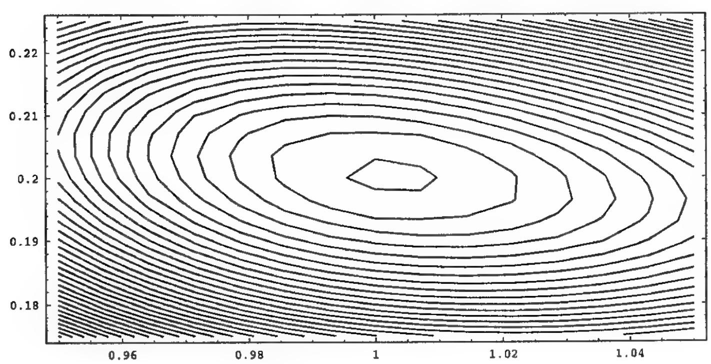
Finally we make one further observation on curve fitting based on our experience in constructing these examples. The large sample sizes and noise levels in the examples severely tested all the algorithms used for finding closest points on the curve: their odd failure for particular configurations in turn induced large errors in the calculated fit. For example, the closest point can be characterised as the root of a quartic, but care must be taken in constructing this quartic otherwise it becomes degenerate or close to degenerate in certain configurations. This in turn means the wrong root is chosen as the closest point, producing an outlier in the distribution of orthogonal distances. As is usual in least squares regression, such outliers can strongly influence the computed fit: when it occured this effect overrode the bias.



(a) 5000 noisy points.

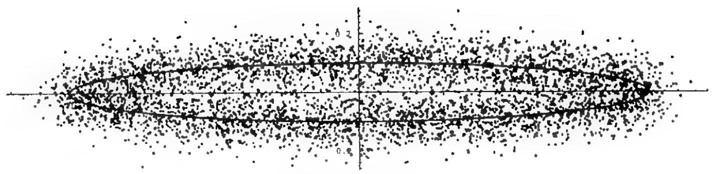


(b) Contours of the sum of squared orthogonal distances.

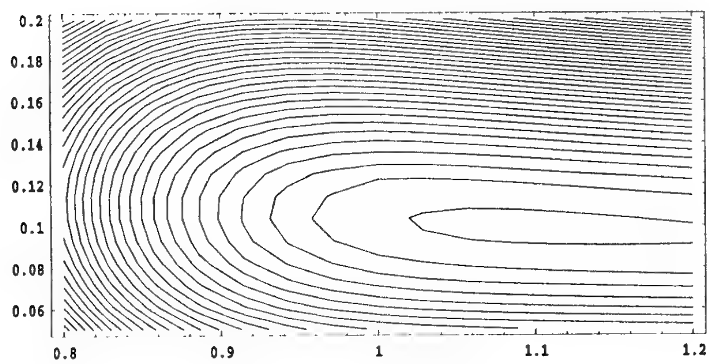


(c) Contours of the second order approximation to the log-likelihood function.

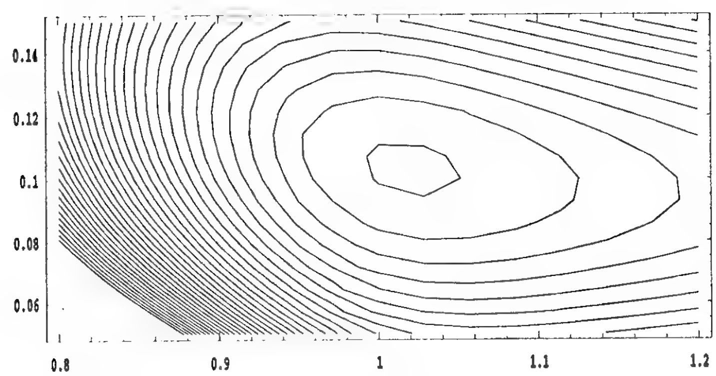
Figure 3: The ellipse with $a = 1$ *and* $b = 0.2$ *is sampled uniformly along its arc length and then iid zero-mean Gaussian noise with* $\sigma = 0.05$ *is added to the* x *and* y *coordinate of the sampled point to produce the 5000 noisy points shown in (a) above. The contours of the sum of squared orthogonal distances (2) and the second-order approximation to the log-likelihood function (13) are shown in (b) and (c) respectively.*



(a) 5000 noisy points.



(b) Contours of the sum of squared orthogonal distances.



(c) Contours of the second order approximation to the log-likelihood function.

Figure 4: The ellipse with $a = 1$ and $b = 0.1$ is sampled uniformly along its arc length and then iid zero-mean Gaussian noise with $\sigma = 0.05$ is added to the x and y coordinate of the sampled point to produce the 5000 noisy points shown in (a) above. The contours of the sum of squared orthogonal distances (2) and the second-order approximation to the log-likelihood function (13) are shown in (b) and (c) respectively.

7 Conclusions

The analysis here confirms intuition and experience by showing that orthogonal distance does indeed give biased fits when the noise level is significantly large compared to the radius of curvature. Moreover, while maximum likelihood estimates based on nonlinear models are generally biased [2], the bias here is qualitatively different in that it is independent of the number of data points. This makes orthogonal distance regression an inconsistent estimator.

This has been noted before [6], but the analysis proposed here to account for it appears to be novel. In [6] the curve is defined implicitly and bias is reduced by making a second order correction to the constraint rather than the penalty function. Minimal justification is given for this step, and the resulting expression is not easily expressed in the terminology used here (in practice it appears to result in a minimisation similar to that in (15) but without the term involving the length of the curve).

The major advantage of the analysis here is that it highlights the likely cause of bias and inconsistency. The maximum likelihood estimation problem (1) defined by the standard model does not have a fixed set of parameters: instead the number of unknown parameters $\mathbf{y}_n$ increases proportionally with the data. Thus standard asymptotic results on convergence of the estimator do not apply. By extending the basic model we are able to integrate out the unknown $\mathbf{y}_n$ at the cost of an extra assumption: that the samples are uniformly distributed along the true curve. The result is a more complete description of the problem that includes all its stochastic components and requires the modeller to make explicit assumptions about each. This focus on a statistical model for curve fitting has the added benefit in that it provides a framework for subsequent processing: for example, the problem of merging separate curve segments can be veiwed as an exercise in hypothesis testing (the tests may be based on comparisons of the distributions of the residuals from separate fits vis à vis the residuals of a joint fit).

There are a number of obvious issues and extensions that we do not have the space to address in detail here, but that deserve mention. First, in many problems the curve length $|c|$ cannot be expressed in closed form but must be approximated. Ideally this would be computed directly by some form of interpolation of the calculated closest points $\mathbf{x}_n^*$. Second, in practice the distribution of samples along the true curve is not likely to be uniform, especially if the curve is infinite. In this case the model could be extended to include the unknown density and this could be estimated as part of the likelihood maximisation. (This would also remove the need to estimate $|c|$.) Third, the data points and their perturbations from the true curve are often higly correlated: this occurs whenever a curve is fitted to an extracted edge or skeleton. In this case the model should be extended to include a model for correlated noise and estimation of any undetermined parameters in this: if it isn't, in our experience the fits produced are still reasonable, but post processing activities such as curve merging [16] fail because the statistical assumption on which they are based are violated.

Finally the most important question is in what circumstances is the fit obtained by minimising (13) likely to produce better results than standard orthogonal distance regression? While we do not have a definitive answer, three observations on it are worth noting. First, in practice there no significant difference in the computational complexity of the

two minimisations. The computational cost of both is dominated by finding the closest point on the curve to each data point: both metrics can then be simply expressed in terms of this distance and te local curvature. Second, the new metric proposed here rests on the assumption that the underlying true samples are distributed uniformly along the curve. The assumption is perfectly reasonable, especially if there is no a priori information to the contrary, but if it is not true then the results in section 5 the bias is likely to be same for both methods. Finally the analysis of circle fitting suggests that bias will dominate ucertainty in the fit produced by orthogonal distance regression if the local density $d(s)$ of observations per unit arc length along the curve exceeds $r(s)/\sigma^2$, where $r(s)$ is the local radius of curvature.

References

1. P. T. Boggs, R. H. Byrd, J. E. Rogers, R. B. Schnabel, "User's reference guide for ODRPACK version 2.01 software for weighted orthogonal distance regression", Applied and Computational Mathematics Division, National Institute of Standards and Technology, U.S. Department of Commerce, NISTIR 92–4834, 1992.

2. M. J. Box, "Bias in nonlinear estimation", Journal of the Royal Statistical Society Series B, 33, 1971, pp. 171–201.

3. G. E. P. Box, G. C. Tiao, Bayesian Inference in Statistical Analysis, John Wiley & Sons, 1973.

4. X. Cao, N. Shrikhande, G. Hu, "Approximate orthogonal distance regression method for fitting quadratic surfaces to range data", Pattern Recognition Letters, 15, 1994, pp. 781–796.

5. G. F. Carrier, M. Krook, C. E. Pearson, Functions of a Complex Variable, McGraw-Hill, New York, 1966.

6. W. A. Fuller, Measurement Error Models, John Wiley & Sons, 1987.

7. I. S. Gradshteyn, I. M. Ryzhik, Table of Integrals Series and Products, Academic Press, New York, 1993.

8. M. Gulliksson, I. Söderkvist, "Surface fitting and parameter estimation with nonlinear least squares", Optimization Methods and Software, 5, 1995, pp. 247–269.

9. R. Hermann, Differential Geometry and the Calculus of Variations, Academic Press, New York, NY, 1968.

10. K. Kanatani, Geometric Computation for Machine Vision, Clarendon Press, Oxford, 1993.

11. K. Kanatani, "Statistical bias of conic fitting and renormalization", IEEE Transactions on Pattern Analysis and Machine Intelligence, 16(3), 1994, pp. 320–326.

12. K. Kanatani, Statistical Optimization for Geometric Computation: Theory and Practice, Elsevier Science, North-Holland, 1996.

13. K. Kanatani, Y. Kanazawa, "Optimal Curve Fitting and Reliability Evaluation", (preprint: `www.ail.cs.gunma-u.ac.jp/~kanatani/curve.ps.gz`), 1997.

14. N. Redding, "The autoscaling of oblique ionograms", Research Report DSTO-RR-0074, DSTO Electronics and Surveillance Research Laboratory, Salisbury, South Australia, Australia, 1996.

15. N. J. Redding, "Image understanding of oblique ionograms: the autoscaling problem", Proceedings of Fourth Australian and New Zealand Conference on Intelligent Information Systems, Adelaide, South Australia, November, 1996, pp. 155–160.

16. N. J. Redding, G. N. Newsam, "A merging procedure for connecting fitted implicit polynomials for features", Proceedings of Fourth Australian and New Zealand Conference on Intelligent Information Systems, Adelaide, South Australia, November, 1996, pp. 299–303.

17. N. J. Redding, "Implicit polynomials, orthogonal distance regression, and the closest point on a curve", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(2), 2000, pp. 191–199.

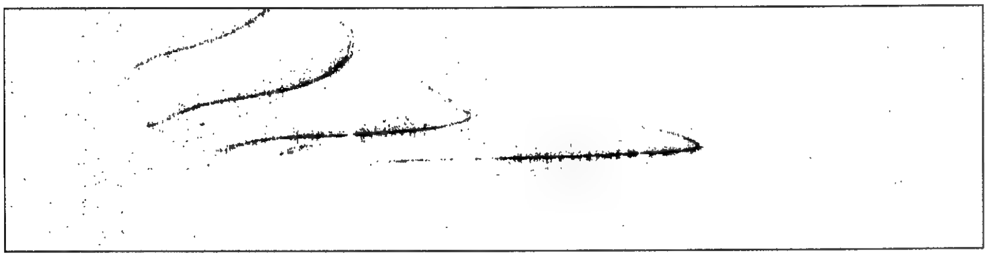


Figure A1: A typical ionogram. Frequency is plotted along the x axis and signal travel-time along the y-axis. The grey-level indicates the strength of the received signal.

Appendix A: Autoscaling Ionograms

High frequency (HF) radio communications over long distances rely on reflections from the ionosphere to bounce the wave from transmitter to receiver around the curve of the earth. The capacity of this indirect channel is a (complicated) function of transmission frequency and the state of the ionosphere, which changes continuously. Therefore communications systems and over-the-horizon (OTH) radar systems that make use of ionospheric reflection require up-to-date information on the state of the ionosphere. This information can be gleaned from *ionograms*: these are visual displays of suitably chosen test signals. Figure A1 shows a typical ionogram: the horizontal axis indicates the frequency of the transmitted signal, the vertical axis indicates the time taken for the signal to reach the receiver and the grey-level indicates the intensity of the received signal.

For the information contained in figure A1 to be useful for HF communications and OTH radar, is has to be *scaled*, *i.e.*, it has to be interpreted by a trained ionospheric physicist. This involves identifying the various curves or *traces* in the image along with their features, and associating these with paths taken by the signal and the atmospheric structure on those paths (refraction by the ionosphere varies with frequency plus a signal path may include more than one reflection between transmitter and receiver). This information can be used both to select optimal operating frequenices for communications or radar systems and to infer the structure of the ionosphere. Ionospheric sounding networks are now recording ionograms over a multitude of paths at intervals of minutes, however, and the resulting data set far outstrips the supply of physicists available to interpret it. Hence there is increasing need for an *autoscaling* system that can automatically interpret recorded ionograms.

Development of a prototype autoscaling system [14, 15, 16, 17] threw up a number of significant curve fitting and post-processing problems which in turn motivated the work presented here. An autoscaling system must be able to identify traces and fit them by a simple parametric model that can then be used for purposes such as classification or tracking of traces. While orthogonal distance regression is an obvious and useful starting point for such a fitting procedure, various features of the autoscaling problem require a more sophisticated model of how data is generated and, consequently, of the optimal fitting process.

For example, any trace fitting routine must give an accurate estimate of the *nose* of

the dominant trace, as this gives the optimal frequency for transmission (transmissions at lower frequencies will be corrupted by multi-pathing; transmissions at higher frequencies will not be reflected by the ionosphere). Likewise the particular measurements selected for use in trace fitting are often correlated or have other significant statistical structure. Finally, fitted traces must often be further processed, *e.g.*, separate segments of the same trace must be merged. All of these activities require a good statistical model of trace formation that can then be used to construct well-founded fitting and merging routines. This paper aims to provide a statistical setting for orthogonal distance regression that enables it to be extended to meet these demands.

Appendix B: Moments of the Conditional Distributions Associated with Circles

This appendix derives various closed form expressions and asymptotic expansions for the moments of the conditional distributions of noisy observations of points uniformly distributed along circles.

Let c be a circle of radius r centred on the origin. We wish to calculate the moments of the conditional distribution $p(\mathbf{x}|c,\sigma)$ of (16), in particular the moments $E[\rho^k]$ where $\rho = \|\mathbf{x}\|$. From (16),

$$\begin{aligned} E[\rho^k] &= \frac{1}{\sigma^2}\int_0^\infty \rho^k e^{-(\rho-r)^2/2\sigma^2} e^{-r\rho/\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right)\rho\, d\rho \\ &= \frac{e^{-r^2/2\sigma^2}}{\sigma^2}\int_0^\infty \rho^{k+1} e^{-\rho^2/2\sigma^2} I_0\left(\frac{r\rho}{\sigma^2}\right) d\rho \\ &= (2\sigma^2)^{k/2} e^{-r^2/2\sigma^2}\Gamma\left(\frac{k}{2}+1\right)\, {}_1F_1\left(\frac{k}{2}+1, 1, \frac{r^2}{2\sigma^2}\right) \end{aligned} \qquad \text{(B1)}$$

from 6.631.1 of [7]. Trivially, $E[\rho^0] = 1$. To compute $E[\rho]$, note that 9.215.3 and 9.212 of [7] give

$$\begin{aligned} {}_1F_1(1/2,1,z) &= e^{z/2} I_0(z/2), \\ \frac{d\,{}_1F_1}{dz}(1/2,1,z) &= \frac{1}{2}\,{}_1F_1(3/2,2,z), \\ z\,{}_1F_1(3/2,2,z) &= {}_1F_1(3/2,1,z) - {}_1F_1(1/2,1,z). \end{aligned}$$

Therefore,

$$\begin{aligned} {}_1F_1(3/2,1,z) &= 2z\frac{d}{dz}\,{}_1F_1(\frac{1}{2},1,z) + {}_1F_1(\frac{1}{2},1,z) \\ &= z e^{z/2} I_1(z/2) + (z+1)e^{z/2} I_0(z/2). \end{aligned}$$

Substituting this back into (B1) gives

$$E[\rho] = \sqrt{\frac{\pi}{2}}\sigma e^{-r^2/4\sigma^2}\left[\left(\frac{r^2}{2\sigma^2}+1\right) I_0\left(\frac{r^2}{4\sigma^2}\right) + \frac{r^2}{2\sigma^2} I_1\left(\frac{r^2}{4\sigma^2}\right)\right].$$

We are especially interested in the asymptotic region in which $r/\sigma \gg 1$: in this case substituting in the appropriate asymptotic expansions for I_0 and I_1 gives

$$\begin{aligned} E[\rho] &\sim \frac{\sigma^2}{r}(\frac{r^2}{2\sigma^2}+1)\left[1+\frac{1}{2}\left(\frac{\sigma}{r}\right)^2+\frac{9}{8}\left(\frac{\sigma}{r}\right)^4\right] + \frac{r}{2}\left[1-\frac{3}{2}\left(\frac{\sigma}{r}\right)^2-\frac{15}{8}\left(\frac{\sigma}{r}\right)^4\right] \\ &\sim r\left[1+\frac{1}{2}\left(\frac{\sigma}{r}\right)^2+\frac{1}{8}\left(\frac{\sigma}{r}\right)^4+O\left(\frac{\sigma}{r}\right)^6\right]. \end{aligned} \qquad \text{(B2)}$$

A closed form for $E[\rho^2]$ follows from noting that 9.215.1 and 9.212.4 of [7] gives

$$\begin{aligned} {}_1F_1(1,1,z) &= e^z \\ {}_1F_1(2,1,z) &= (z+1)\,{}_1F_1(1,1,z) = (z+1)e^z. \end{aligned}$$

Substituting this back into (B1) gives

$$E[\rho^2] = r^2 + 2\sigma^2. \tag{B3}$$

Next $E[\rho^4]$ follows from noting that applying 9.212.4 of [7] again gives

$${}_1F_1(3, 1, z) = \frac{(z^2 + 4z + 2)}{2} e^z$$

so that

$$E[\rho^4] = r^4 + 8r^2\sigma^2 + 8\sigma^4. \tag{B4}$$

Appendix C: Fitting a Surface to Points in Space

This section extends the results in the body of the paper on fitting a curve through points in a plane to the problem of fitting a d–1 dimensional surface S through a set of points $\{\mathbf{x}_n\}_{n=1}^N$ in $\mathbb{R}^d$. In particular it derives a second order approximation to the log-likelihood function based on a second order asymptotic approximation to the integrals defining the conditional probabilities $p(\mathbf{x}_n|S,\sigma)$. It is not as easy to work with the quadratic surfaces that are natural local second order approximations to an arbitrary surface as it was with circles. Therefore we instead develop a local approximation to S about the closest point $\mathbf{x}_n^*$ to $\mathbf{x}_n$ as a parabolic surface whose form and orientation is determined by the principal radii of curvature of S at $\mathbf{x}_n^*$. The approach can be readily extended to the more general problem of fitting an d–k dimensional surface S through points in $\mathbb{R}^d$, but for simplicity we shall limit ourselves to the case $k=1$.

In particular, let S be a given d–1 dimensional surface in $\mathbb{R}^d$ and let $\mathbf{x}$ be a noisy observation of S. As before we assume that the probability that $\mathbf{x}$ is a perturbation of a point $\mathbf{y}$ on S is given by

$$p(\mathbf{x}|\mathbf{y},S,\sigma)=\frac{1}{[2\pi\sigma^2]^{d/2}}e^{-\|\mathbf{x}-\mathbf{y}\|^2/2\sigma^2}d\mathbf{x}.$$

Thus the a priori assumption that $\mathbf{y}$ will be uniformly distributed over S gives

$$p(\mathbf{x}|S,\sigma)=\frac{1}{[2\pi\sigma^2]^{d/2}\,|S|}\int_{\mathbf{y}(s)\in S}e^{-\|\mathbf{x}-\mathbf{y}(s)\|^2/2\sigma^2}ds \tag{C1}$$

where $|S|$ is the area of S, and ds is an element of surface area.

We now set up a local approximation to the integral in (C1). To do so, we first define a local coordinate system as follows. Let $\mathbf{x}_n^*$ be the closest point to $\mathbf{x}_n$ on S: $\mathbf{x}_n^*$ will be the origin for the coordinate system. Next let T_n be the tangent plane to S at $\mathbf{x}_n^*$. The directions associated with the principal curvatures of S at $\mathbf{x}_n^*$ form an orthogonal basis for T_n [9]; let the unit vectors in these directions be $\mathbf{e}_1,\dots,\mathbf{e}_{d-1}$. Then, since $\mathbf{x}_n-\mathbf{x}_n^*$ is orthogonal to T_n, we may take the final coordinate direction $\mathbf{e}_d$ to be the unit vector in the direction $\mathbf{x}_n-\mathbf{x}_n^*$.

As before, let $d_n\equiv\|\mathbf{x}_n-\mathbf{x}_n^*\|$. Rather than interpreting d_n as a signed distance and using this to describe the local geometry, however, we now instead take the signs of the curvatures $\kappa_1,\dots,\kappa_{d-1}$ to be with respect to the direction $\mathbf{e}_d$. Thus in this coordinate system the point $\mathbf{x}_n$ has coordinates $(0,\dots,0,d_n)$, and the surface S can be locally represented as the following parabola over the tangent plane

$$S\sim\{(y_1,\dots,y_{d-1},y_d(y_1,\dots,y_{d-1}))\ :\ (y_1,\dots,y_{d-1})\in T_n\}$$

where

$$y_d(y_1,\dots,y_{d-1})=\frac{1}{2}\sum_{i=1}^{d-1}\kappa_i y_i^2.$$

Likewise we can approximate the surface element ds by the elemental product $dy_1 \dots dy_{d-1}$. Thus (C1) can be approximated by

$$p(\mathbf{x}_n|S,\sigma) \sim \frac{1}{[2\pi\sigma^2]^{d/2}\,|S|} \int_{\mathbb{R}^{d-1}} e^{-\|\mathbf{x}_n-\mathbf{y}(s)\|^2/2\sigma^2} dy_1 \dots dy_{d-1}.$$

Next expanding the distance $\|\mathbf{x}_n - \mathbf{y}(s)\|^2/2\sigma^2$ as

$$\begin{aligned}
\frac{1}{2\sigma^2}\|\mathbf{x}_n - \mathbf{y}(s)\|^2 &= \frac{1}{2\sigma^2}\left[\sum_{i=1}^{d-1} y_i^2 + (y_d - d_n)^2\right] \\
&= \frac{1}{2\sigma^2}\left[\sum_{i=1}^{d-1} y_i^2 + \left(\frac{1}{2}\sum_{i=1}^{d-1}\kappa_i y_i^2 - d_n\right)^2\right] \\
&= \frac{1}{2}\left[\left(\sum_{i=1}^{d-1}\frac{\sigma\kappa_i}{2}\left(\frac{y_i}{\sigma}\right)^2\right)^2 + \sum_{i=1}^{d-1}(1-\kappa_i d_n)\left(\frac{y_i}{\sigma}\right)^2 + \left(\frac{d_n}{\sigma}\right)^2\right]
\end{aligned}$$

and making the change of variables $z_i = y_i/\sigma$ gives

$$p(\mathbf{x}_n|S,\sigma) \sim \frac{e^{-d_n^2/2\sigma^2}}{[2\pi]^{d/2}\,\sigma|S|} \int_{\mathbb{R}^{d-1}} \exp\left[-\frac{1}{8}\left(\sum_{i=1}^{d-1}\sigma\kappa_i z_i^2\right)^2\right] \cdot \exp\left[-\frac{1}{2}\sum_{i=1}^{d-1}(1-\kappa_i d_n) z_i^2\right] dz_1 \dots dz_{d-1}.$$

Under the assumptions that

$$\sigma\kappa_i \ll 1, \qquad d_n\kappa_i \ll 1, \qquad \text{and} \qquad \frac{d_n}{\sigma} \sim 1$$

then an asymptotic expansion of the integral above in each variable gives to first order

$$p(\mathbf{x}_n|S,\sigma) \sim \frac{e^{-d_n^2/2\sigma^2}}{\sqrt{2\pi}\,|S|\sigma} \prod_{i=1}^{d-1}(1-\kappa_i d_n)^{-1/2}.$$

Thus the likelihood function is

$$\begin{aligned}
l(\mathbf{x}_n|S,\sigma) &\equiv -\ln p(\mathbf{x}_n|S,\sigma) \qquad\qquad (C2)\\
&\sim \ln\left[\sqrt{2\pi}\,|S|\sigma\right] + \frac{d_n^2}{2\sigma^2} + \frac{1}{2}\sum_{i=1}^{d-1}\ln(1-\kappa_i d_n) \\
&\sim \ln\left[\sqrt{2\pi}\,|S|\sigma\right] + \frac{d_n^2}{2\sigma^2} - \frac{d_n}{2}\sum_{i=1}^{d-1}\kappa_i \\
&\sim \ln\left[\sqrt{2\pi}\,|S|\sigma\right] + \frac{d_n^2}{2\sigma^2} - \frac{(d-1)d_n\bar{\kappa}_n}{2}
\end{aligned}$$

where $\bar{\kappa}_n \equiv \frac{1}{d-1}\sum_{i=1}^{d-1}\kappa_i$ is the mean curvature of S at $\mathbf{x}_n^*$. Finally, given N independent observations the log-likelihood function is approximated by

$$l(\mathbf{x}_1,\dots,\mathbf{x}_N : S,\sigma) \sim \sum_{n=1}^{N}\left[\frac{d_n^2}{2\sigma^2} - \frac{(d-1)\bar{\kappa}_n d_n}{2}\right] + N\left[\ln|S| + \ln\sqrt{2\pi}\sigma\right].$$

This approximation is an obvious extension of (13) after making the simple replacement of the curvatures κ_i by the corresponding radii of curvature $r_i = 1/\kappa_i$. Again, if the variance is unknown it can be eliminated by observing that at the minimum $\sigma^2 = \frac{1}{N}\sum_{n=1}^{N} d_n^2$. Substituting in this value, exponentiating and squaring gives that an approximation to the maximum likelihood estimate of S will be the minimiser of

$$e_{2,d}(\mathbf{x}_1, \ldots, \mathbf{x}_N : S) \equiv |S|^2 \cdot \left(\frac{1}{N}\sum_{n=1}^{N} d_n^2\right) \cdot \exp\left[-\frac{d-1}{N}\sum_{n=1}^{N} \bar{\kappa}_n d_n\right].$$

DISTRIBUTION LIST

Fitting the Most Likely Curve through Noisy Data

Garry N. Newsam and Nicholas J. Redding

	Copy Number
DEFENCE ORGANISATION	
Task Sponsor	
Director, DIGO	1
S&T Program	
Chief Defence Scientist	2
FAS Science Policy	
AS Science Corporate Management	
Director General Science Policy Development	
Counsellor, Defence Science, London	3
Counsellor, Defence Science, Washington	Doc Data Sht
Scientific Adviser to MRDC, Thailand	Doc Data Sht
Scientific Adviser Joint	4
Navy Scientific Adviser	Doc Data Sht
Scientific Adviser, Army	Doc Data Sht
Air Force Scientific Adviser	5
Director Trials	6
Aeronautical and Maritime Research Laboratory	
Director, Aeronautical and Maritime Research Laboratory	7
Electronics and Surveillance Research Laboratory	
Director, Electronics and Surveillance Research Laboratory	Doc Data Sht
Chief, Surveillance Systems Division	8
Research Leader, Imagery Systems	9
Head, Image Analysis & Exploitation	10
Head, Imaging Radar Systems	11
Guy Blucher	12
Dr David Crisp	13
David I. Kettler	14
Dr Tim Payne	15
Mark Preiss	16
Dr Nicholas J. Redding	17 to 22
Dr Nick J. S. Stacy	23
Bob Whatmough	24

DSTO Research Library and Archives	
Library Fishermans Bend	Doc Data Sht
Library Maribyrnong	Doc Data Sht
Library Edinburgh	25 to 26
Australian Archives	27
Library, MOD, Pyrmont	Doc Data Sht
Library, MOD, HMAS Stirling	28
US Defense Technical Information Center	29 to 30
UK Defence Research Information Centre	31 to 32
Canada Defence Scientific Information Service	33
NZ Defence Information Centre	34
National Library of Australia	35
Capability Systems Staff	
Director General Maritime Development	Doc Data Sht
Director General Land Development	36
Director General Aerospace Development	Doc Data Sht
Navy	
SOSCI, Surface Combatants FEG Wartime Division, Maritime HQ, Garden Island	Doc Data Sht
Army	
ABCA Standardisation Officer, Puckapunyal	37 to 40
SO(Science), DJFHQ(L), MILPO, Enoggera, Queensland 4057	Doc Data Sht
NPOC QWG Engineer NBCD Combat Development Wing, Puckapunyal	41
Air Force	
Knowledge Staff	
Director General Command, Control, Communications and Computers	Doc Data Sht
Director General Intelligence, Surveillance, Reconnaissance and Electronic Warfare	Doc Data Sht
Director General Defence Knowledge Improvement Team	Doc Data Sht
Intelligence Program	
DGSTA, Defence Intelligence Organisation	42
Manager, Information Centre, Defence Intelligence Organisation	43
Acquisitions Program	
Corporate Support Program	
Library Manager, DLS-Canberra	44

UNIVERSITIES AND COLLEGES

Australian Defence Force Academy Library	45
Head of Aerospace and Mechanical Engineering, ADFA	46
Deakin University Library, Serials Section (M List)	47
Hargrave Library, Monash University	Doc Data Sht
Librarian, Flinders University	48

OTHER ORGANISATIONS

NASA (Canberra)	49
AusInfo	50
State Library of South Australia	51

ABSTRACTING AND INFORMATION ORGANISATIONS

Library, Chemical Abstracts Reference Service	52
Engineering Societies Library, US	53
Materials Information, Cambridge Scientific Abstracts, US	54
Documents Librarian, The Center for Research Libraries, US	55

INFORMATION EXCHANGE AGREEMENT PARTNERS

Acquisitions Unit, Science Reference and Information Service, UK	56
Library – Exchange Desk, National Institute of Standards and Technology, US	57
National Aerospace Laboratory, Japan	58
National Aerospace Laboratory, Netherlands	59

SPARES

DSTO Salisbury Research Library	60 to 64

Total number of copies:	64

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA				1. CAVEAT/PRIVACY MARKING	
2. TITLE Fitting the Most Likely Curve through Noisy Data			3. SECURITY CLASSIFICATION Document (U) Title (U) Abstract (U)		
4. AUTHORS Garry N. Newsam and Nicholas J. Redding			5. CORPORATE AUTHOR Electronics and Surveillance Research Laboratory PO Box 1500 Edinburgh, South Australia, Australia 5111		
6a. DSTO NUMBER DSTO–RR–0242	6b. AR NUMBER 012–374		6c. TYPE OF REPORT Research Report	7. DOCUMENT DATE May, 2002	
8. FILE NUMBER B 9505–21–70	9. TASK NUMBER INT 99/184	10. SPONSOR DDIGO	11. No OF PAGES 31		12. No OF REFS 17
13. URL OF ELECTRONIC VERSION http://www.dsto.defence.gov.au/corporate/reports/DSTO–RR–0242.pdf			14. RELEASE AUTHORITY Chief, Surveillance Systems Division		
15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT *Approved For Public Release* OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, SALISBURY, SOUTH AUSTRALIA 5108					
16. DELIBERATE ANNOUNCEMENT No Limitations					
17. CITATION IN OTHER DOCUMENTS No Limitations					
18. DEFTEST DESCRIPTORS orthogonal distance regression fitting ellipse fitting					

19. ABSTRACT

At present the preferred method for fitting a general curve through scattered data points in the plane is *orthogonal distance regression*, *i.e.,* by minimising the sum of squares of the distances from each data point to its nearest neighbour on the curve. While generally producing good fits, in theory orthogonal distance regression can be both biased and inconsistent: in practice this manifest itself in overfitting of convex curves or underfitting of corners. The paper postulates this occurs because orthogonal distance regression is based on an incomplete stochastic model of the problem. It therefore presents an extension of the standard model that takes into accounts both the noisy measurement of points on the curve and their underlying distribution along the curve. It then derives the likelihood function of a given curve being observed under this model. Although this cannot be evaluated exactly for anything other than the simplest curves, it lends itself naturally to asymptotic approximation. Orthogonal distance regression corresponds to a first order approximation to the maximum likelihood estimator in this model: the paper also derives a second order approximation, which turns out to be a simple modification of the least squares penalty that includes a contribution from the curvature at the closest point. Analytical and numerical examples are presented to demonstrate the improvement achieved using the higher order estimator.